

基于 RAD 高通量测序探讨中国 85 种 杜鹃花属植物的分类

李云飞^{1,2,3*}, 李世明^{1,2,3*}, 金鑫^{1,2}, 程书^{2,3}, 王松波^{2,3}, 侯军亮^{1,2,3},
刘家劲^{1,2,3,4}, 段肖霞^{1,2,3,4}, 马宏⁵, 马永鹏^{6**}, 张耕耘^{1,2,3,4**}

(1. 深圳华大生命科学研究院, 广东 深圳 518083; 2. 深圳市华大农业应用研究院, 广东 深圳 518120; 3. 农业基因组学国家重点实验室, 广东 深圳 518083; 4. 农业部农业基因组学重点实验室, 广东 深圳 518083; 5. 中国林业科学研究院资源昆虫研究所, 云南 昆明 650224; 6. 云南省极小种群野生植物综合保护重点实验室, 中国科学院昆明植物研究所, 云南 昆明 650201)

摘要: [目的] 探讨 RAD-seq 技术在中国杜鹃花复杂类群分类与物种界定方面的优势。[方法] 本研究对杜鹃花属 85 个种进行了 RAD-seq 测序, 评估了数据的基本特征; 同时, 以马缨杜鹃基因组作为参考, 获得高质量 SNP 位点, 并进一步通过 ADMIXTURE、PCA、GCTA 及 FastTree 软件对其进行基因分型、聚类与系统树构建。[结果] 本研究中 85 份材料比对至马缨杜鹃的平均比对率 87.52%, 在参考基因组上的平均覆盖度为 5.21%, 最终获得了 620 371 个 SNP 位点。基于 PCA、Structure 聚类及系统发育树的分析结果, 本研究支持目前在亚属水平(杜鹃亚属、常绿杜鹃亚属、马银花亚属、羊躑躅亚属、映山红亚属和长蕊杜鹃亚属)的形态学分类处理, 并突出了鳞片在杜鹃属植物分类中的重要作用。[结论] RAD-seq 数据产生的大量 SNP 位点可区分杜鹃花亚属和组内的大部分物种, 说明其在复杂类群分类与物种界定方面具有可行性。

关键词: RAD 测序; 杜鹃花; 单核苷酸多态性; 资源利用

中图分类号: S718.3

文献标识码: A

文章编号: 1001-1498(2019)03-0001-08

杜鹃花是杜鹃花科(Ericaceae)杜鹃属(*Rhododendron* L.)植物的统称,是世界著名观赏植物,中国十大传统名花、云南省八大名花之一。杜鹃花在全世界有 1 000 余种,广泛分布于亚洲、北美洲和欧洲。中国是野生杜鹃花资源最丰富的国家之一。自《中国植物志》、*Floral of China* 出版以后,近年来大量新种被陆续描述^[1]和归并^[2]。据统计,当前我国现存的杜鹃花属植物已经超过 600 种。除宁夏和新疆等干旱荒漠地带外,中国南北各省均有杜鹃花分布^[3],其中,西南横断山脉片区野生杜鹃花资源最密集,是杜鹃花的多度中心和多样化中心。

我国横断山区杜鹃花的繁衍方式属于植物快速的辐射分化^[4]与种间频繁的自然杂交^[5-8],严重制

约了该属植物的科学分类研究与准确鉴定,进而减缓了我国在杜鹃花资源挖掘与利用方面的进程。近年来,我国不同学者尝试用一代测序的方法(如 DNA barcoding 片段(ITS、rbcL、matK 和 psbA-trnH)测序法)分别在杜鹃花亚组、组、亚属和整个属水平上做了探索,但是这些片段在物种、亚组、甚至组间的区分能力较弱^[9-11]。

近年来,随着新一代测序技术的不断发展,测序成本不断下降,以单核苷酸多态性(SNP)为代表的新型标记逐渐被应用到植物群体遗传学研究中。SNP 位点具有数量多、分布密度高等特点,而且 SNP 位点的开发不需要进行 PCR 和电泳等实验工作,可以大规模高度自动化地完成。

收稿日期: 2018-09-19 修回日期: 2019-03-18

基金项目: 深圳市科技计划项目(JCYJ20150831201123287); 深圳作物分子设计聚合育种工程实验室提升项目(深发改[2015]946号); 广东省农作物核心资源开发应用企业重点实验室(2011A091000047)

* 同等贡献作者: 李云飞,男,主要研究方向:植物群体遗传学;李世明,主要研究方向:生物信息学。

** 通讯作者: 马永鹏,研究员。研究方向:重要观赏植物类群的保护与资源利用。E-mail: mayongpeng@mail.kib.ac.cn; 张耕耘,研究员。研究方向:基因组学与植物育种。E-mail: zhanggengyun@genomics.cn

本研究采用了 RAD-seq 技术(在二代测序基础上发展起来的一种 DNA 测序新技术)来开发单核苷酸多态性 SNP 位点,该技术利用限制性内切酶对基因组进行酶切,产生一定大小的片段,构建测序文库,对酶切后产生的 RAD 标记进行高通量测序。RAD-seq 能够较均匀的覆盖整个基因组,进而可以获得成千上万的 SNP 位点,以期发现 RAD-seq 技术在中国杜鹃花复杂类群分类与物种界定方面的优势。

1 材料与方法

1.1 实验材料

杜鹃花样本:用于 RAD-seq 的杜鹃花种类及其所属的亚组、组和亚属见表 1。本实验测序的样本共 85 种,隶属于 6 个亚属、8 个组和 22 个亚组。

1.2 实验流程

每个个体摘取新鲜叶片 0.5 g,用 CTAB 法^[12]分别提取样品的基因组 DNA 并进行 DNA 质量检测,对质量合格的 DNA 进行下一步实验。

1.3 RAD 文库构建

参考 Miller 等^[13]的文库构建方法,分以下步骤:① 限制性内切酶 *Taq* I 酶切;② 连接 P1 接头;③ DNA 随机打断片断化;④ 目的片段回收与末端修复;⑤ 连接 P2 接头;⑥ RAD 片段富集;⑦ 上机测序。

1.4 测序 reads 的过滤

根据识别标签序列得到每个个体的测序 reads,先使用 trimmomatic-0.36^[14]对测序 reads 进行过滤,设置具体参数如下,过滤参数为:SLIDINGWINDOW:5:20 LEADING:5 TRAILING:5 MINLEN:50。过滤标准为:两端质量低于 5 的碱基进行切除,并以 5 bp 为窗口进行滑动过滤,对平均质量低于 20 的窗口进行切除。

1.5 比对与变异检测

用 BWA 软件^[15] mem 模块将过滤后的个体 clean reads 比对到马缨花杜鹃参考基因组序列上。样本比对率反映的是样本测序数据与参考基因组的相似性,覆盖深度和覆盖度能够直接反映测序数据的均一性与参考序列的同源性。

比对完成后,经过 GATK 软件^[16]进行变异检测,并获得群体变异集文件(VCF 格式)。最后对原

始基因型数据基本特性进行分析后,对数据进行过滤:过滤参数为缺失率小于或等于 0.2、杂合率小于或等于 0.2、最小等位基因频率(MAF)大于或等于 0.05,最终得到高质量的基因型数据。

1.6 物种聚类分析

在获得高质量的标记数据以后,使用 ADMIXTURE^[17]、GCTA^[18]、FastTree^[19] 软件进行物种的聚类分析。

2 结果与分析

2.1 数据产出

对 85 个种的杜鹃花样本进行 RAD 简化基因组测序,共获得测序数据量 41.56 G,平均每个个体 488.93 Mb。使用 trimmomatic-0.36 软件对拆分好的测序数据进行过滤后得到 230.44 M 条长度为 50~144 bp 的 clean reads,平均长度为 137.9 bp。数据量为 31.79 G,平均每个个体 374.03 Mb。另外,数据的 Q20、Q30 也分别从过滤前的 93.3、84.7 提高到过滤后的 97.7、92.6,有效的保证了后续分析的准确性。

使用 BWA^[15] 软件 mem 的模块,把每个个体的 clean data 比对到马缨杜鹃基因组上,并使用 samtools 对比对结果进行统计。比对结果为 85 个个体的 reads 比对到马缨杜鹃全基因组的平均比对率为 88.52%,覆盖参考基因组的区域平均为 7.81%,而覆盖区域的平均深度为 5.21X(表 1)。比对完成后,利用 GATK 的 Haplotype Caller 模块对比对结果进行变异检测,得到群体的 SNP 位点集,并经过严格过滤,最终得到 620 371 个 SNP。

2.2 群体结构分析

使用 plink 软件随机选择连锁不平衡(LD)小于 0.1,且相邻间隔在 300 kb 以上的 SNP 位点,最后得到一个包含 3 420 个 SNP 位点的标记集。使用 ADMIXTURE 软件对此 SNP 位点集进行群体结构分析,为了确定合适的群体数(K 值),预先假定 K 值为 1~8,分别对每个假设的 K 值进行运算,并使用交叉验证过程确定 K 值,根据软件说明, cross-validation error 值最低时所对应的 K 值为最合适的 K 值。结果显示,当 $K=2$ 时,曲线出现最低值;同时,考虑到样本所归属的分类单元, $K=3$ 和 $K=4$ 也做了 Structure 聚类图。

表 1 杜鹃花样品编号、种名及归属、数据产出、测序深度、覆盖度和马缨杜鹃全基因组的对比率

Table 1 Basic information of *Rhododendron* species sequenced in the present study and characteristics of RAD data including clean reads, clean base, depth, coverage percentage as well as mapping percentage to the whole genome of *R. delavayi*

样品编号 Sample ID	物种名 Species	所属亚属 Subgenus	所属组或亚组 Section/Subsection	高质量序列 Clean reads	高质量碱基 Clean base/Mb	测序深度/倍 Depth	覆盖度 Coverage /%	比对率 Mapping percent/%
1	皱皮杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	3.06	422.50	6.87	7.92	88.92
2	黄花杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	2.95	408.91	6.37	6.67	89.16
3	高尚杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.41	470.15	7.62	8.68	97.20
6	怒江杜鹃	杜鹃亚属	杜鹃组 - 怒江杜鹃亚组	2.37	323.77	4.25	8.52	91.49
8	皱叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.73	236.22	3.59	8.49	89.65
10	绒毛杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 麻花杜鹃亚组	2.03	282.00	6.07	4.54	71.97
11	滇南杜鹃	马银花亚属	长蕊组	1.63	224.91	4.03	5.72	84.58
12	柳条杜鹃	糙叶杜鹃亚属	帚枝杜鹃组	1.90	260.56	3.88	7.69	91.65
13	隐脉杜鹃	杜鹃亚属	杜鹃组 - 有鳞大花亚组	1.80	249.42	4.29	5.94	86.43
14	不凡杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.43	196.60	3.78	6.50	86.60
16	团叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.51	485.10	6.23	10.33	92.51
17	宝兴杜鹃	杜鹃亚属	杜鹃亚属 - 川西杜鹃亚组	2.67	368.33	4.82	6.07	70.37
18	棕背杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	2.21	304.54	4.38	8.85	88.95
20	粗脉杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	3.09	426.61	5.41	10.43	93.86
22	白毛杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	3.16	437.55	5.22	11.22	92.95
23	羊躑躅	羊躑躅亚属	五花药组	1.45	198.02	3.90	5.22	87.93
24	峨眉银叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	2.26	312.03	5.04	8.39	93.87
25	锈红杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	2.83	391.97	3.24	6.27	36.78
26	锦锈杜鹃	映山红亚属	映山红组	1.90	260.77	4.27	6.86	91.34
28	稀果杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 麻花杜鹃亚组	2.63	362.67	5.25	9.21	93.23
32	繁花杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.29	176.32	3.58	6.41	89.77
34	大白杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	2.31	319.98	5.25	8.78	98.90
36	紫花杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	2.74	375.05	4.48	8.97	87.99
37	耳叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 耳叶杜鹃亚组	4.30	595.98	8.40	9.74	95.11
38	美艳橙红杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 火红杜鹃亚组	4.78	657.59	9.31	9.83	96.22
39	美容杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.54	488.29	6.64	9.45	91.66
42	糙叶杜鹃	糙叶杜鹃亚属	糙叶杜鹃组	4.52	624.53	7.30	9.31	90.66
44	长鳞杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 麻花杜鹃亚组	2.46	339.53	4.34	8.12	72.04
45	百合花杜鹃	杜鹃亚属	杜鹃组 - 有鳞大花亚组	3.13	428.51	7.29	5.93	85.79
47	基毛杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	1.28	174.94	3.42	5.59	92.45
48	三花杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	2.42	333.56	5.92	6.05	93.51
50	阔柄杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	2.79	384.11	7.03	7.62	96.35
51	岷江杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.58	218.87	4.28	7.13	96.37
52	陇蜀杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	2.25	308.59	4.96	8.23	92.31
53	四川杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.57	490.86	6.93	9.50	93.46
56	椿年杜鹃	杜鹃亚属	杜鹃组 - 有鳞大花亚组	2.55	353.77	6.11	5.74	85.54
57	鳞腺杜鹃	杜鹃亚属	杜鹃组 - 鳞腺杜鹃亚组	1.20	164.87	3.97	4.55	89.70
58	巴郎杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大理杜鹃亚组	1.86	258.13	5.19	6.90	95.38
59	问客杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	4.44	615.93	6.96	8.93	85.49
60	照山白	杜鹃亚属	杜鹃组 - 照山白亚组	1.19	165.24	3.67	5.01	88.44
61	锈叶杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	2.60	361.13	4.77	8.05	88.90
62	长毛杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	3.75	519.61	5.82	9.90	90.88
63	卧龙杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	1.91	265.56	4.21	7.22	79.40
64	滇隐脉杜鹃	杜鹃亚属	杜鹃组 - 有鳞大花亚组	5.21	719.36	7.88	9.72	91.48
65	大果杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	2.30	316.30	5.04	8.73	96.48
66	碎米花杜鹃	糙叶杜鹃亚属	糙叶杜鹃组	7.40	1 017.40	14.04	7.66	94.16
67	大喇叭杜鹃	杜鹃亚属	杜鹃组 - 有鳞大花亚组	2.69	369.45	6.91	5.65	90.32
68	云上杜鹃	杜鹃亚属	杜鹃组 - 有鳞大花亚组	2.18	301.42	5.26	6.05	94.15
69	桃叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 露珠杜鹃亚组	3.84	528.20	7.61	9.64	95.67

续表 1

样品编号 Sample ID	物种名 Species	所属亚属 Subgenus	所属组或亚组 Section/Subsection	高质量序列 Clean reads	高质量碱基 Clean base/Mb	测序深度/倍 Depth	覆盖度 Coverage / %	比对率 Mapping percent/ %
70	薄叶马银花	马银花亚属	马银花组	1.47	202.87	4.12	5.42	88.56
71	红棕杜鹃	杜鹃亚属	杜鹃组 - 亮鳞杜鹃亚组	4.40	609.12	6.82	9.71	92.83
72	金山杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.25	171.74	3.48	6.80	95.18
73	马缨杜鹃	常绿杜鹃亚属	常绿杜鹃亚组 - 树形杜鹃亚组	2.31	319.85	4.01	11.13	94.64
78	紫玉盘杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 镰果杜鹃亚组	1.97	273.04	4.79	7.67	93.37
79	粘毛杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 粘毛杜鹃亚组	2.28	316.73	4.62	7.80	78.77
80	山育杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	2.89	397.33	4.70	9.58	92.67
81	毛棉杜鹃	马银花亚属	长蕊组	1.71	236.57	4.93	5.10	86.94
84	多鳞杜鹃	杜鹃亚属	杜鹃组 - 三花杜鹃亚组	2.10	290.23	4.70	6.02	78.95
85	云锦杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	2.49	341.82	4.90	9.42	94.10
86	反边杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	2.93	404.22	5.55	8.78	83.92
88	越峰杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.10	428.02	7.11	7.95	91.21
89	碟花杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 露珠杜鹃亚组	1.13	155.92	3.39	6.50	96.98
90	金山杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.81	248.51	3.88	8.57	92.54
91	纯红杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 火红杜鹃亚组	1.83	250.89	3.44	7.52	72.58
92	大钟杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	2.89	397.34	5.17	9.85	89.14
93	腺果杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.59	496.44	6.88	9.74	93.55
94	皱叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	3.34	458.72	5.76	10.07	87.93
95	昆明杜鹃	杜鹃亚属	糙叶杜鹃组	2.29	317.74	5.50	6.59	92.98
96	腋花杜鹃	糙叶杜鹃亚属	腋花杜鹃组	4.84	670.27	11.18	6.60	92.44
97	长蕊杜鹃	马银花亚属	长蕊组	1.15	156.79	3.52	4.85	88.61
98	爆杖花	糙叶杜鹃亚属	糙叶杜鹃组	4.13	566.53	7.21	8.25	89.00
99	喉斑杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 蜜腺杜鹃亚组	4.30	596.80	9.14	8.61	91.00
100	圆头杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 杯毛杜鹃亚组	2.09	289.82	4.74	8.18	92.27
102	美容杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	2.40	331.56	4.69	9.18	90.29
103	大王杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 杯毛杜鹃亚组	3.49	479.40	6.25	9.42	84.73
104	大王杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 杯毛杜鹃亚组	3.50	485.68	7.46	8.14	86.38
105	长粗毛杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 粘毛杜鹃亚组	2.26	309.69	4.55	7.80	79.36
106	巨魁杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大叶杜鹃亚组	2.11	290.94	4.48	8.44	89.65
107	波叶杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	2.19	298.46	4.14	9.39	90.31
108	棉毛房杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 星毛杜鹃亚组	2.87	398.01	5.52	10.01	97.07
109	井冈山杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	3.08	427.25	4.90	8.82	71.76
110	喇叭杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 云锦杜鹃亚组	4.61	637.76	9.71	8.48	89.60
111	凸尖杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 大叶杜鹃亚组	2.00	278.63	4.21	6.51	68.80
112	海绵杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 银叶杜鹃亚组	1.78	245.55	1.65	0.99	4.94
113	夺目杜鹃	常绿杜鹃亚属	常绿杜鹃组 - 杯毛杜鹃亚组	3.74	516.64	7.13	8.23	78.22

从 Structure 聚类图 1 可以看出:当 $K=2$ 时,2 个不同的遗传组成(红色和蓝色种类)分别对应常绿杜鹃亚属(110~73)、蓝色对应羊躑躅亚属(23)、映山红亚属(26)、马银花亚属(11、81、97 和 70);除糙叶杜鹃亚属的腋花杜鹃组(12)外,该亚属其余种类与杜鹃亚属完全聚在一起(13~80)。当 $K=3$ 时,除常绿杜鹃亚属,其余分组都没有变化。常绿杜鹃亚属(110~73)再分成了 2 个组分:即云锦杜鹃亚组种类(110~34)和常绿杜鹃亚属的其余种类(16~73)。当 $K=4$ 时,常绿杜鹃亚属的分组和 $K=3$ 时一致;其它亚属种类:97~67(其中 97、11、81 为马银花亚属;13、45、56、67 为杜鹃亚属有鳞大花

杜鹃亚组);17~57(其中 17 为川西杜鹃亚组,68、64 为有鳞大花杜鹃亚组,60 为照山白杜鹃组,5 为鳞腺杜鹃亚组)。

2.3 主成份分析

主成分分析是研究群体遗传结构的一种重要的多元统计方法。在本研究中,笔者使用 GCTA 软件进行主成分分析。首先,利用包含 620 371 个 SNP 位点的数据集构建所有样本间的亲缘关系矩阵;其次,使用构建好的亲缘关系矩阵计算前 3 个特征值和特征向量;最后,使用 R 脚本图形展示,并根据检索得到样品的亚属信息进行分类。

PCA 聚类结果(图 2)与 Structure 结果相似:85

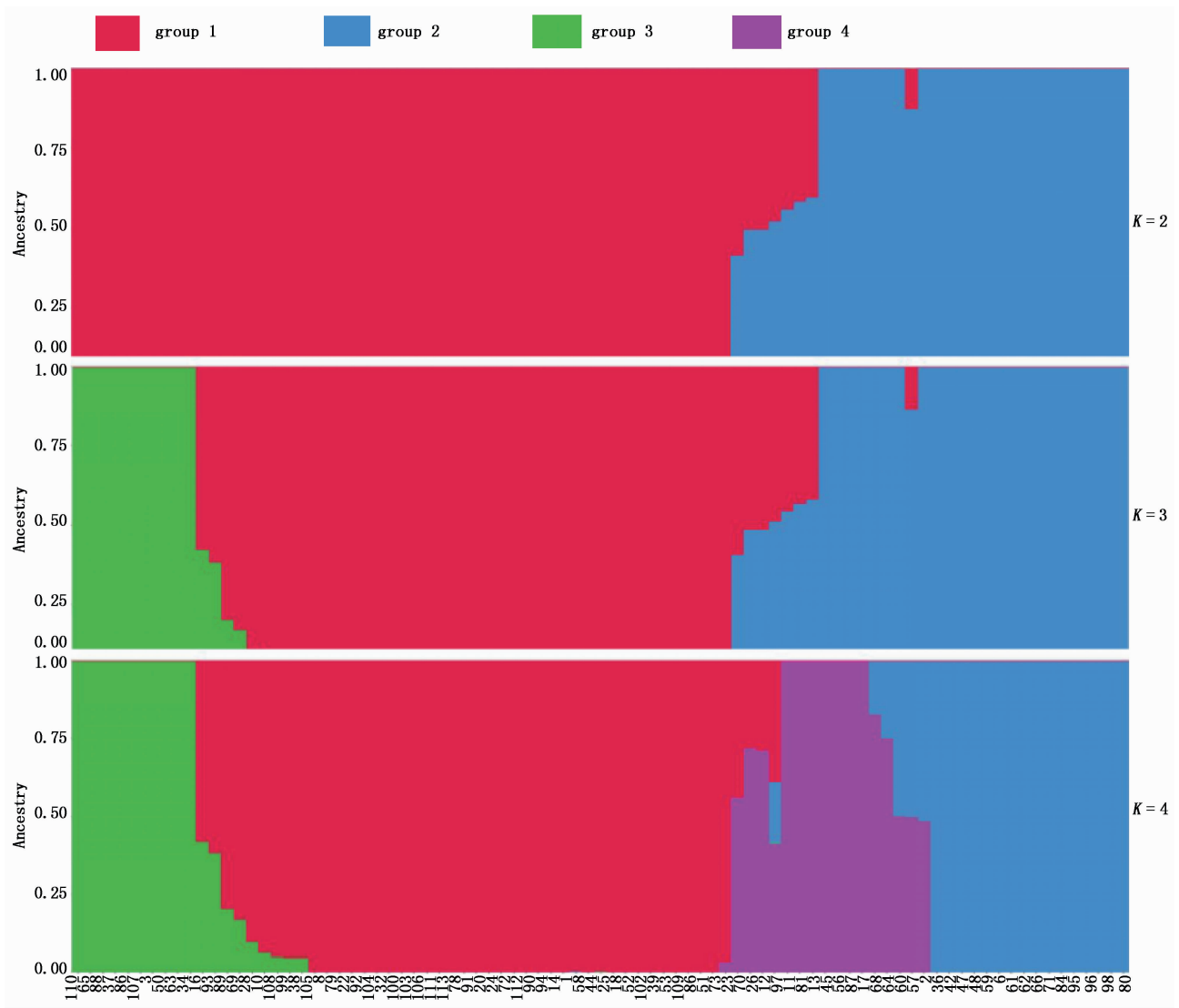


图 1 基于 3 420 个 SNP 位点的 85 个杜鹃花样本的 Structure 聚类图(上:K=2;中:K=3;下 K=4)

Fig.1 Results from Structure analysis of 85 *Rhododendron* species based on 3 420 SNPs

(Top: K=2; Middle: K=3; Bottom: K=4).

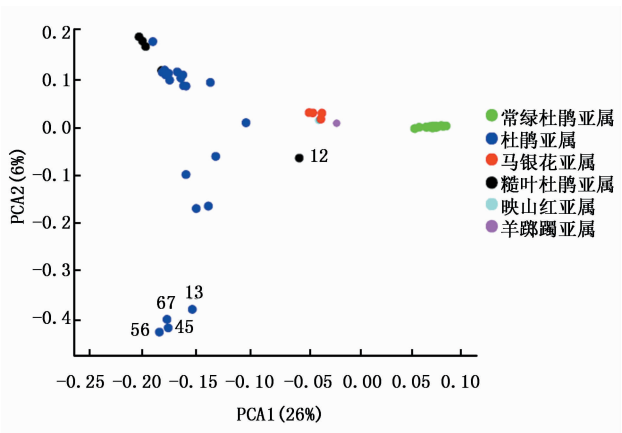


图 2 基于 620 371 个 SNP 位点的 85 个杜鹃花样本的 PCA 聚类图

Fig.2 Results from PCA analysis of 85 *Rhododendron* species based on 620 371 SNPs

个物种的样本分为 5 个聚类:常绿杜鹃亚属(绿色圆点),羊躑躅亚属、马银花亚属和映山红亚属种类聚在一起(蓝绿、浅红与深红色圆点),糙叶杜鹃亚属的腋花杜鹃组(黑色圆点 12),杜鹃亚属的有鳞大花杜鹃亚组种类(蓝色圆点 13、45、56、67),以及杜鹃亚属(其它蓝色圆点)和糙叶杜鹃亚属的其它种类(其它黑色圆点)。

2.4 系统发育树构建

系统发育树是反映群体中每个个体关系的重要形式。在本研究中,为了了解各样本间的亲缘关系,笔者采用 FastTree 软件构建系统发育树,使用 GTR + CAT 模型进行最大似然亲缘关系分析。基于 620 371个 SNP 位点的 85 个种杜鹃花样本的系统发育树(图 3)表明:常绿杜鹃亚属聚为独立一支;马银

红亚属和羊躑躅亚属独立于常绿杜鹃亚属和杜鹃亚属,推测是 2 个大亚属的一个过渡,也进一步暗示常绿杜鹃亚属和杜鹃亚属物种的分化早于马银花亚属、映山红亚属和羊躑躅亚属的分化。

本研究的结果,将为今后的杜鹃花资源深度挖掘与利用提供重要的指导:(1)常绿杜鹃亚属与杜鹃亚属由于分化较早,亲缘关系相对较远。在做杜鹃花杂交育种时,尽量避免用这 2 个亚属内部的物种做人工杂交。这一点在某些远缘杂交的实验中得到证实^[23];(2)鉴于糙叶杜鹃亚属与杜鹃亚属(主要是三花杜鹃亚组种类)的关系较近,可以利用前者较密集的簇生小花序(比如腋花杜鹃)和鲜艳的筒状花(爆仗杜鹃)与花冠较大、适应性较强的云南杜鹃、黄花杜鹃做人工杂交,培育兼具二者优良性状的杜鹃花新品种;(3)有鳞类的杜鹃和无鳞类杜鹃杂交授粉可以尝试以马银花亚属种类和有鳞大花杜鹃亚组种类为杂交材料,前者耐热、花期较早;而后者喜寒、花期较晚,可以培育出一批适合我国西南海拔较低区域生长的杜鹃花品种;(4)从商业化规模生产考虑,优先推荐映山红种类与马银花亚属种类做杂交选育。前者的主要优势在于很多种类耐热,且容易做扦插扩繁;而后者观赏价值较高,且很多种类花香浓郁(比如长蕊杜鹃、滇南杜鹃和毛棉杜鹃等)。二者杂交很容易培育花香、色艳、耐热且极易实现无性大规模扩繁的新品种。

4 结论

本研究对杜鹃花属 85 个种进行了 RAD-seq 测序,评估了数据的基本特征;同时,以马缨杜鹃基因组作为参考,获得高质量 SNP 位点;基于这些 SNP 位点对杜鹃花属植物进行基因分型、聚类与系统树构建。研究结果支持目前在亚属水平的形态学分类处理,并突出了鳞片在杜鹃属植物分类中的重要作用。

参考文献:

- [1] Cai L, Neilsen J, Dao ZL, *et al.* *Rhododendron longipedicellatum* (Ericaceae), a new species from Southeastern Yunnan, China [J]. *Phytotaxa*, 2016, 282 (4): 296–300.
- [2] Marczewski T, Ma Y P, Zhang X M, *et al.* Why is population information crucial for taxonomy? A case study employing a hybrid swarm in *Rhododendron* and related described varieties [J]. *AoB Plants*, 2016, 8: plw070.
- [3] Ma Y P, Neilsen J, Chamberlin D, *et al.* The conservation of *Rhododendrons* is of greater urgency than has been previously acknowledged in China [J]. *Biodiversity and Conservation*, 2014, 23(12):

3149–3154.

- [4] Milne R I, Davies C, Prickett R, *et al.* Phylogeny of *Rhododendron* subgenus *Hymenanthes* based on chloroplast DNA markers: between lineage hybridization during adaptive radiation [J]. *Plant Systematics and Evolution*, 2010, 285 (3–4): 233–244.
- [5] Zha H G, Milne R I, Sun H. Asymmetric hybridization in *Rhododendron agastum*: a hybrid taxon comprising mainly F_1 s in Yunnan, China [J]. *Annals of Botany*, 2010, 105 (1): 89–100.
- [6] Ma Y P, Milne R I, Zhang C Q, *et al.* Unusual patterns of hybridization involving a narrow endemic *Rhododendron* species in Yunnan (Ericaceae), China [J]. *American Journal of Botany*, 2010, 97 (10): 1749–1757.
- [7] Ma Y P, Xie W J, Sun WB, *et al.* Strong reproductive isolation despite occasional hybridization between a widely distributed and a narrow endemic *Rhododendron* species [J]. *Scientific Reports*, 2016, 6(1): 19146.
- [8] Zhang J L, Zhang C Q, Gao L M, *et al.* Natural hybridization origin of *Rhododendron agastum* (Ericaceae) in Yunnan, China: inferred from morphological and molecular evidence [J]. *Journal of Plant Research*, 2007, 120 (3): 457–463.
- [9] Gao L M, Li D Z, Zhang C Q, *et al.* Infrageneric and sectional relationships in the genus *Rhododendron* (Ericaceae) inferred from ITS sequence data [J]. *Journal of Systematics and Evolution*, 2002, 44 (11): 1351–1356.
- [10] 刘仁林. 中国杜鹃属映山红亚属植物系统分类研究 [D]. 南昌: 江西林业大学, 2007.
- [11] Yan L J, Liu J, Michae M, *et al.* DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains [J]. *Molecular Ecology Resources*, 2015, 15(4): 932–944.
- [12] Doyle J J, Doyle J L. Genomic plant DNA preparation from fresh tissue-CTAB method [J]. *Phytochem Bull*, 1987, 19 (11): 11–15.
- [13] Miller M R, Dunham J P, Amores A, *et al.* Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers [J]. *Genome Research*, 2007, 17 (2): 240–248.
- [14] Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, 30 (15): 2114–2120.
- [15] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2010, 26 (5): 589–595.
- [16] McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data [J]. *Genome research*, 2010, 20 (9): 1297–1303.
- [17] Patterson N, Moorjani P, Luo Y, *et al.* Ancient admixture in human history [J]. *Genetics*, 2012, 192 (3): 1065–1093.
- [18] Yang J, Lee S H, Goddard M E, *et al.* GCTA: a tool for genome-wide complex trait analysis [J]. *American Journal of Human Genetics*, 2011, 88 (1): 76–82.
- [19] Price M N, Dehal P S, Arkin A P. FastTree 2-approximately maxi-

mum-likelihood trees for large alignments [J]. PloS one, 2010, 5 (3): e9490.

[20] 耿玉英, 中国杜鹃花属植物[M]. 上海:上海科学技术出版社,2014.

[21] 闵天禄, 方瑞征. 杜鹃属的系统发育与进化[J]. 云南植物研究, 1990, 12 (4): 352 – 365.

[22] Yan L J, Gao L M, Li D Z. Molecular evidence for natural hybridization between *Rhododendron spiciferum* and *R. spinuliferum* (Ericaceae) [J]. Journal of Systematics and Evolution, 2013, 51 (4): 426 – 434.

[23] 庄 平. 32 种杜鹃花属植物亚属间杂交的可育性研究[J]. 广西植物, 2018, DOI: 10.11931/guihaia.gxzw201709009.

Phylogenomic Analysis of 85 *Rhododendron* Species in China
Based on RAD Sequencing

LI Yun-fei^{1,2,3}, LI Shi-ming^{1,2,3}, JIN Xin^{1,2}, CHENG Shu^{2,3}, WANG Song-bo^{2,3}, HOU Jun-liang^{1,2,3},
LIU Jia-jin^{1,2,3,4}, DUAN Xiao-xia^{1,2,3,4}, MA Hong⁵, MA Yong-peng⁶, ZHANG Geng-yun^{1,2,3,4}

(1. BGI Institute of Life Sciences, BGI-Shenzhen, Shenzhen 518083, Guangdong, China; 2. BGI Institute of Applied Agriculture, BGI-Shenzhen, Shenzhen 518120, Guangdong, China; 3. State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, Guangdong, China; 4. Key Laboratory of Genomics, Ministry of Agriculture, BGI-Shenzhen, Shenzhen 518120, Guangdong, China; 5. Research Institute of Resources Insects, Chinese Academy of Forestry, Kunming 650224, Yunnan, China; 6. Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China)

Abstract: [Objective] This study aims at examining the application of RAD-seq in classifications of the genus *Rhododendron*. [Method] 85 species of *Rhododendron* were sequenced using dd-RAD and the basic characteristics data obtained were evaluated. Then, the whole genome sequence of *Rh. delavayi* were used as the reference genome and the SNP loci were effectively mapped and obtained. These SNP loci were used to construct the phylogenomic relationship among groups employed by ADMIXTURE, PCA, GCTA and FastTree software. [Result] On average, 87.52% of reads were mapped to the *Rh. delavayi* genome and the average coverage to the genome was 5.21%. After cleaning and filtering these data, a total of 620 371 SNPs were obtained. Preliminary analysis of these observed SNP loci by ADMIXTURE, PCA and FastTree revealed that the species grouping was corresponding to subgenus division from Floral of China, though different subsections with subgenus were not clearly differentiated. Moreover, the scale underside leaf was emphasized as the key character to distinguishing subgenus with genus *Rhododendron*. [Conclusion] The study proves the potential application of RAD-seq in classifications of the large genus *Rhododendron* and the results will provide insights into resources exploration and utilization of *Rhododendrons* in China.

Keywords: RAD-seq;Rhododendron;SNPs;resources utilization

(责任编辑:张 研)